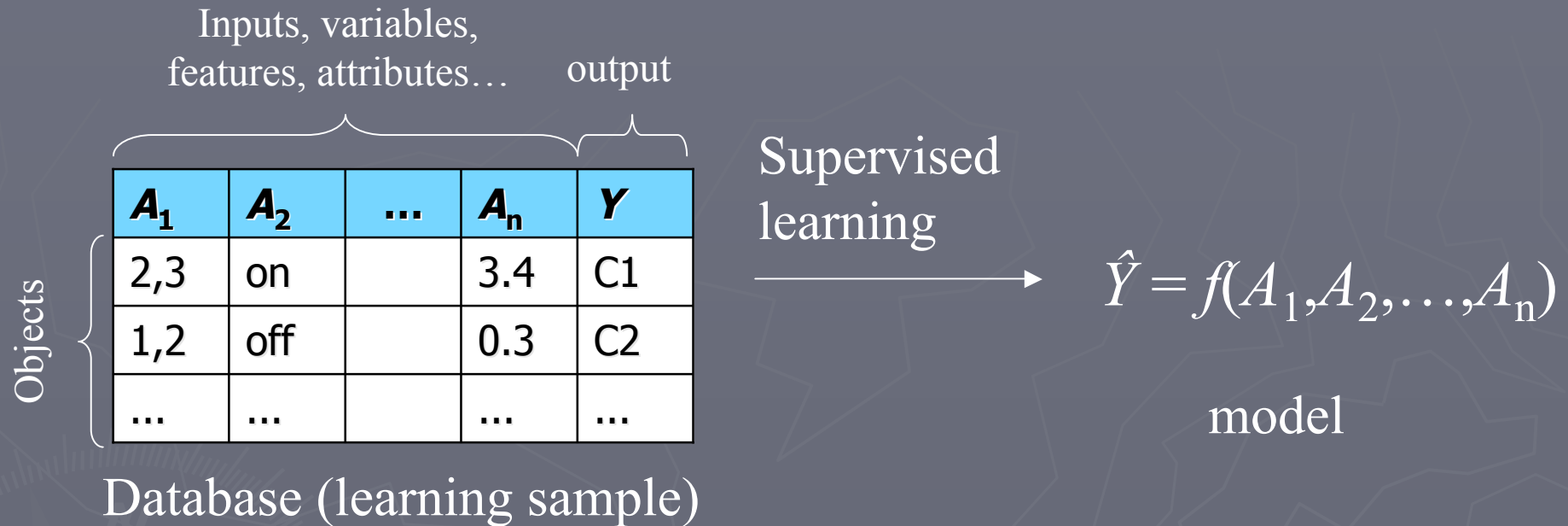


# Apprentissage supervisé: Arbres de décision

AIA - 2006 - Leçon 2

# Generic supervised learning problem



- ▶ Goal: from the database, find a function  $f$  of the inputs that approximates at best the output
- ▶ Symbolic output  $\rightarrow$  classification problem,  
Numerical output  $\rightarrow$  regression problem

# Some (non-biological) applications

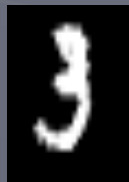
- ▶ Bank credit assignment
  - Decide to accept a credit or not based on past clients

checkingaccount	duration	purpose	amount	savings	yearsemployed	age	Debtor
0<=...<200 DM	48	radiotv	5951	...<100 DM	1<...<4	22	bad
...<0 DM	6	radiotv	1169	unknown	...>7	67	good
no	12	education	2096	...<100 DM	4<...<7	49	good
...<0 DM	42	furniture	7882	...<100 DM	4<...<7	45	good
...<0 DM	24	newcar	4870	...<100 DM	1<...<4	53	bad
no	36	education	9055	unknown	1<...<4	35	good
no	24	furniture	2835	500<...<1000 DM	...>7	53	good
0<=...<200 DM	36	usedcar	6948	...<100 DM	1<...<4	35	good
no	12	radiotv	3059	...>1000 DM	4<...<7	61	good
0<=...<200 DM	30	newcar	5234	...<100 DM	unemployed	28	bad
0<=...<200 DM	12	newcar	1295	...<100 DM	...<1	25	bad
...<0 DM	48	business	4308	...<100 DM	...<1	24	bad
0<=...<200 DM	12	radiotv	1567	...<100 DM	1<...<4	22	good

# Some (non-biological) applications

## ▶ Image classification

- Handwritten character recognition



3

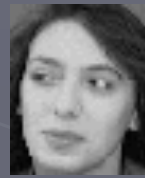


5

- Face recognition



Roger

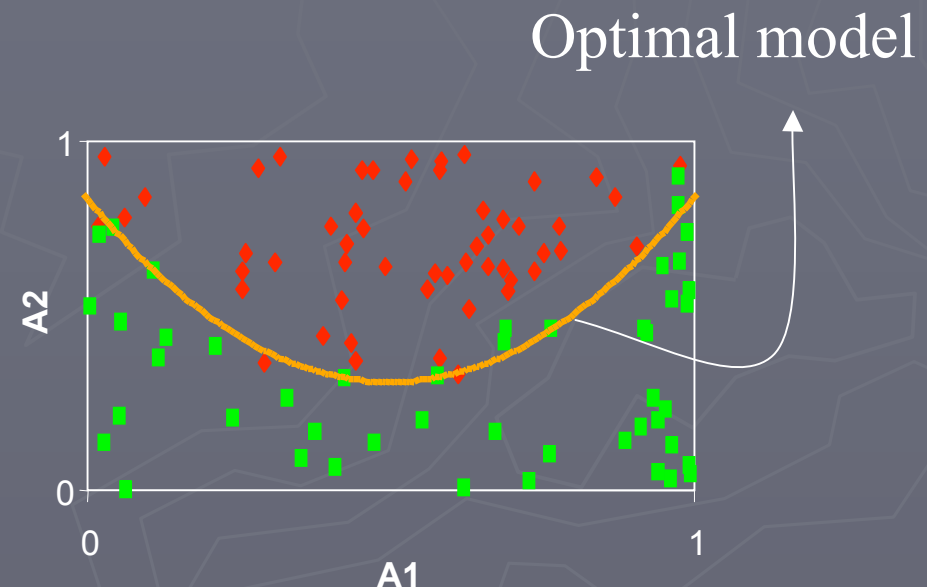


Anna

# Illustrative problem

- ▶ 2 numerical input variables  $A1$  and  $A2$ , discrete output  $C$
- ▶ Learning sample:

$A1$	$A2$	$C$
0.1	0.23	good
0.5	0.6	bad
...	...	...
0.24	0.12	good



- ▶ Goal: find a model that classify at best *new* cases for which  $A1$  and  $A2$  are known
- ▶ Error of the optimal model is called the *residual error*

# Evaluation of learning algorithms

- ▶ Learning algorithms are evaluated according to three main criteria:
  - Accuracy:
    - ▶ Measured by the generalization error (on an independent sample)
  - Efficiency:
    - ▶ Computing times and scalability for learning and testing
  - Interpretability:
    - ▶ Comprehension bought by the model about the input-output relationship
- ▶ There is usually a tradeoff between these criteria

# Two popular learning algorithms

## ▶ Decision trees:

- Very fast and interpretable. Easy to use.
- Often less accurate than neural networks.

## ▶ Artificial neural networks:

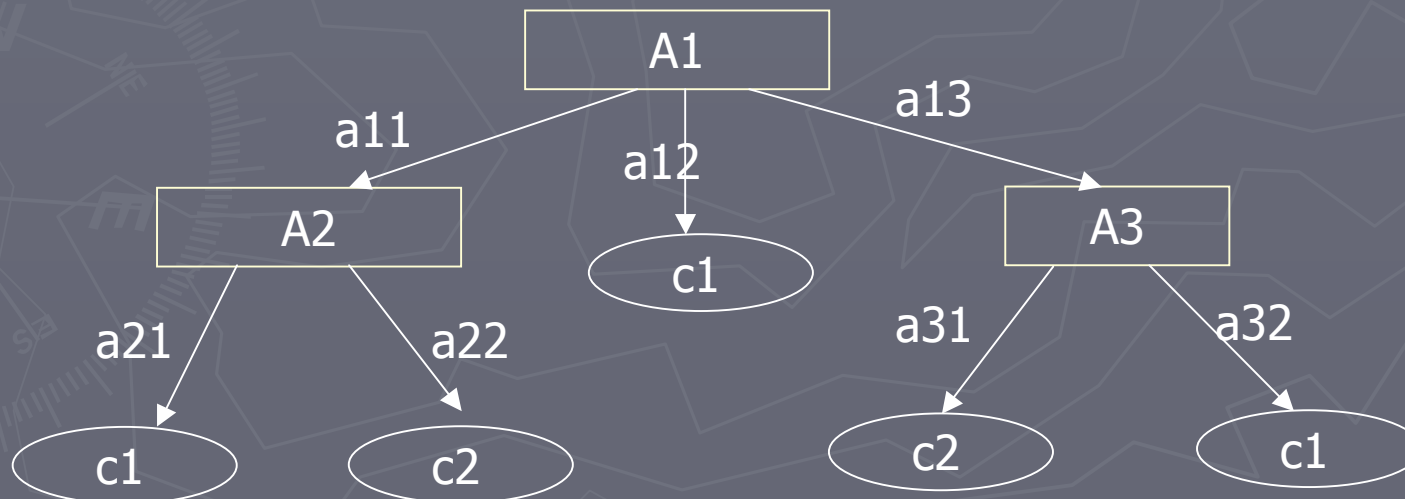
- Often very accurate.
- Very slow and limited interpretability. Difficult to use.

# Decision (classification) trees

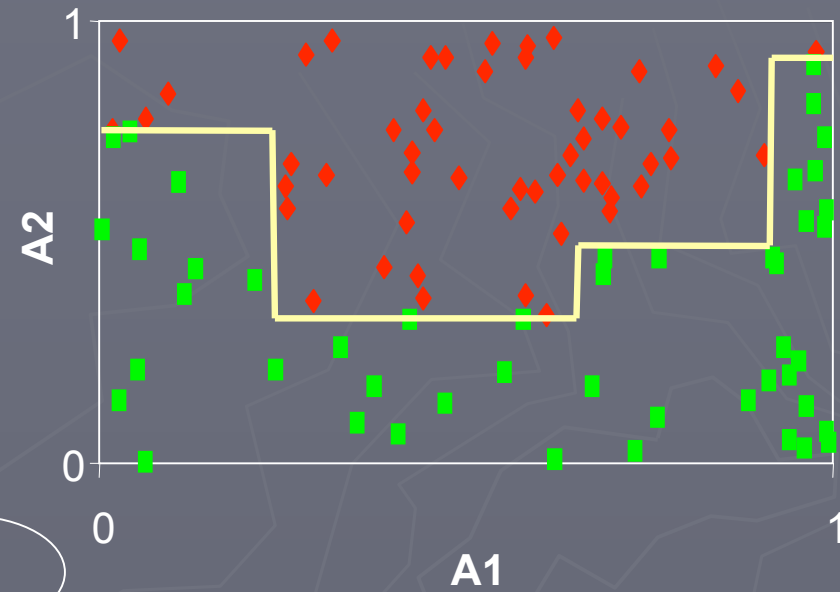
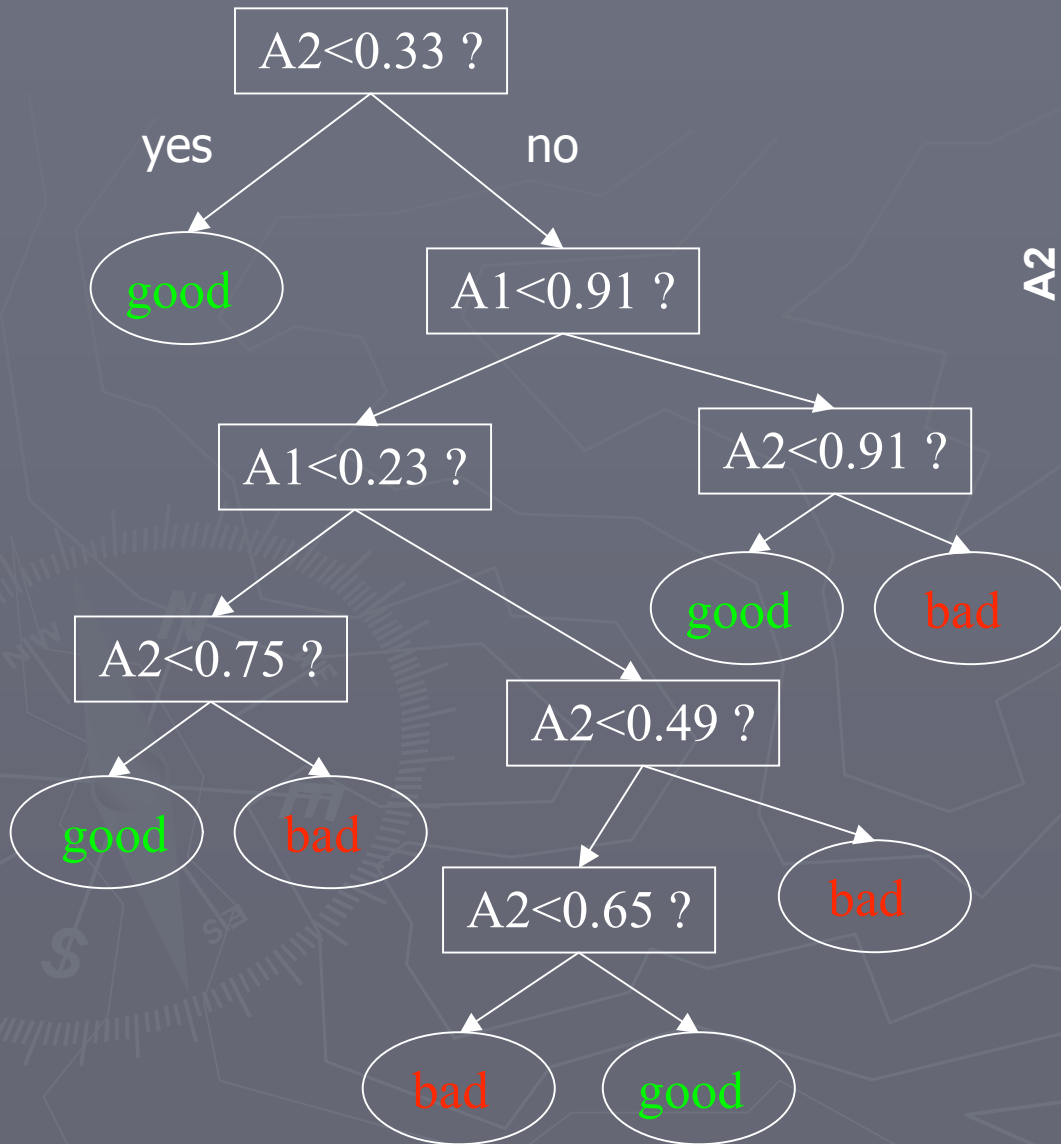
- ▶ A learning algorithm that can handle:
  - Classification problems (binary or multi-valued)
  - Attributes may be discrete (binary or multi-valued) or continuous.
- ▶ Classification trees were invented twice:
  - By statisticians: CART (Breiman et al.)
  - By the AI community: ID3, C4.5 (Quinlan et al.)

# Hypothesis space

- ▶ A decision tree is a tree where:
  - Each **interior node** tests an attribute
  - Each **branch** corresponds to an attribute value
  - Each **leaf** node is labelled with a class



# Illustration

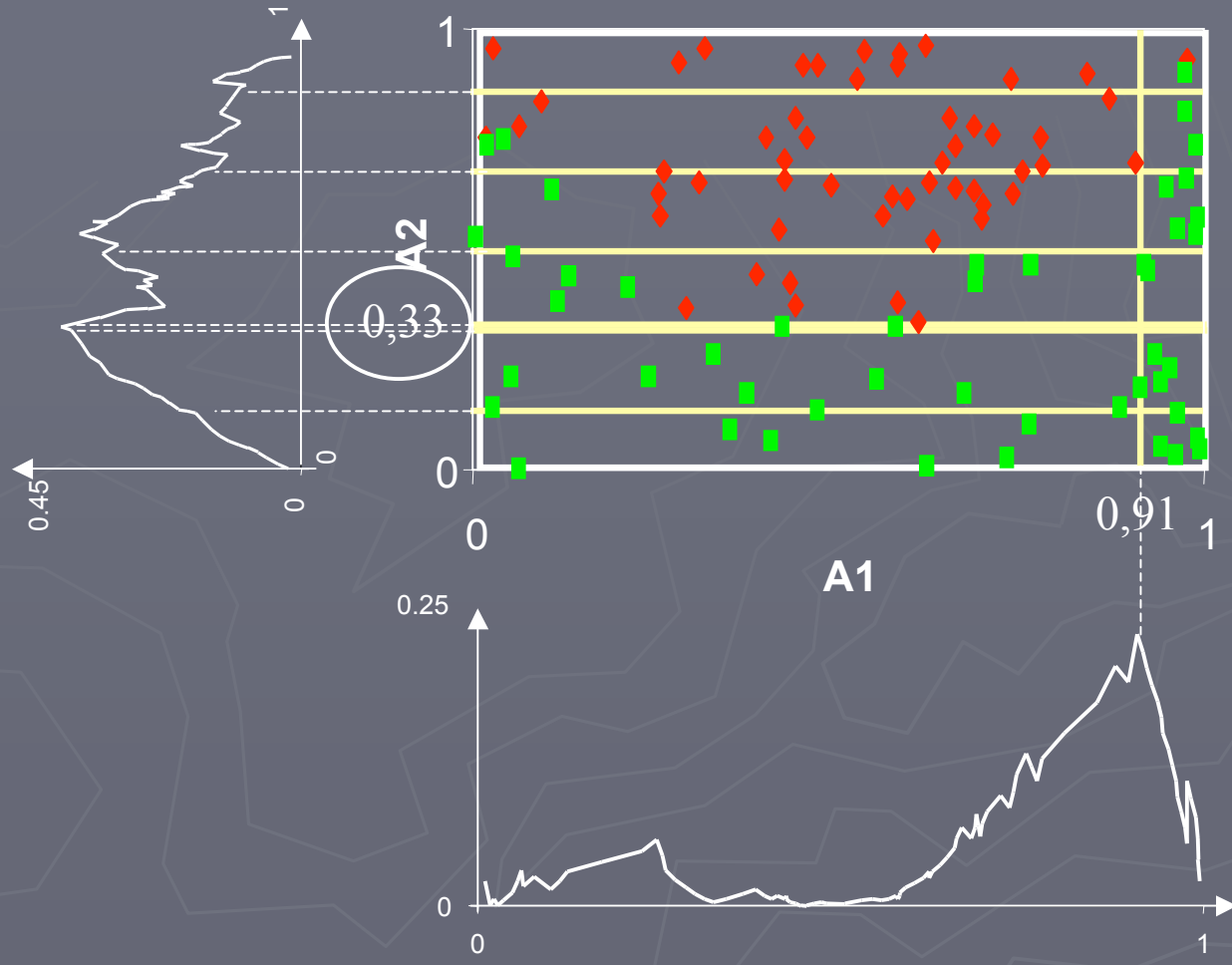


# Learning a tree

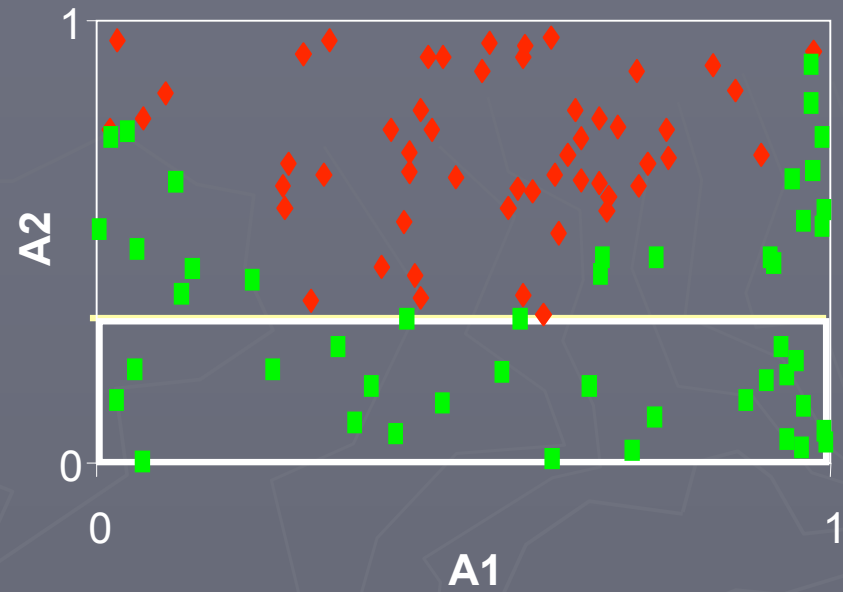
- ▶ Growing: from  $S_1$ 
  - From top to bottom
  - At each step:
    - ▶ Choose a terminal node
    - ▶ Search for the best attribute and the best test
    - ▶ Split the node
- ▶ Pruning: from  $S_2$ 
  - From bottom to top
  - At each step:
    - ▶ Choose a test at the bottom of the tree
    - ▶ Prune the subtree and evaluate the new tree on  $S_2$

# Tree growing

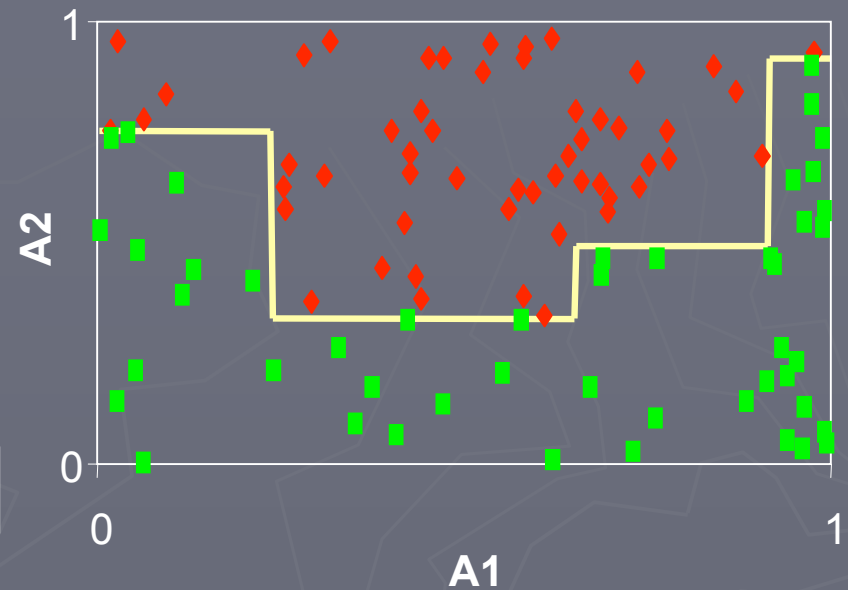
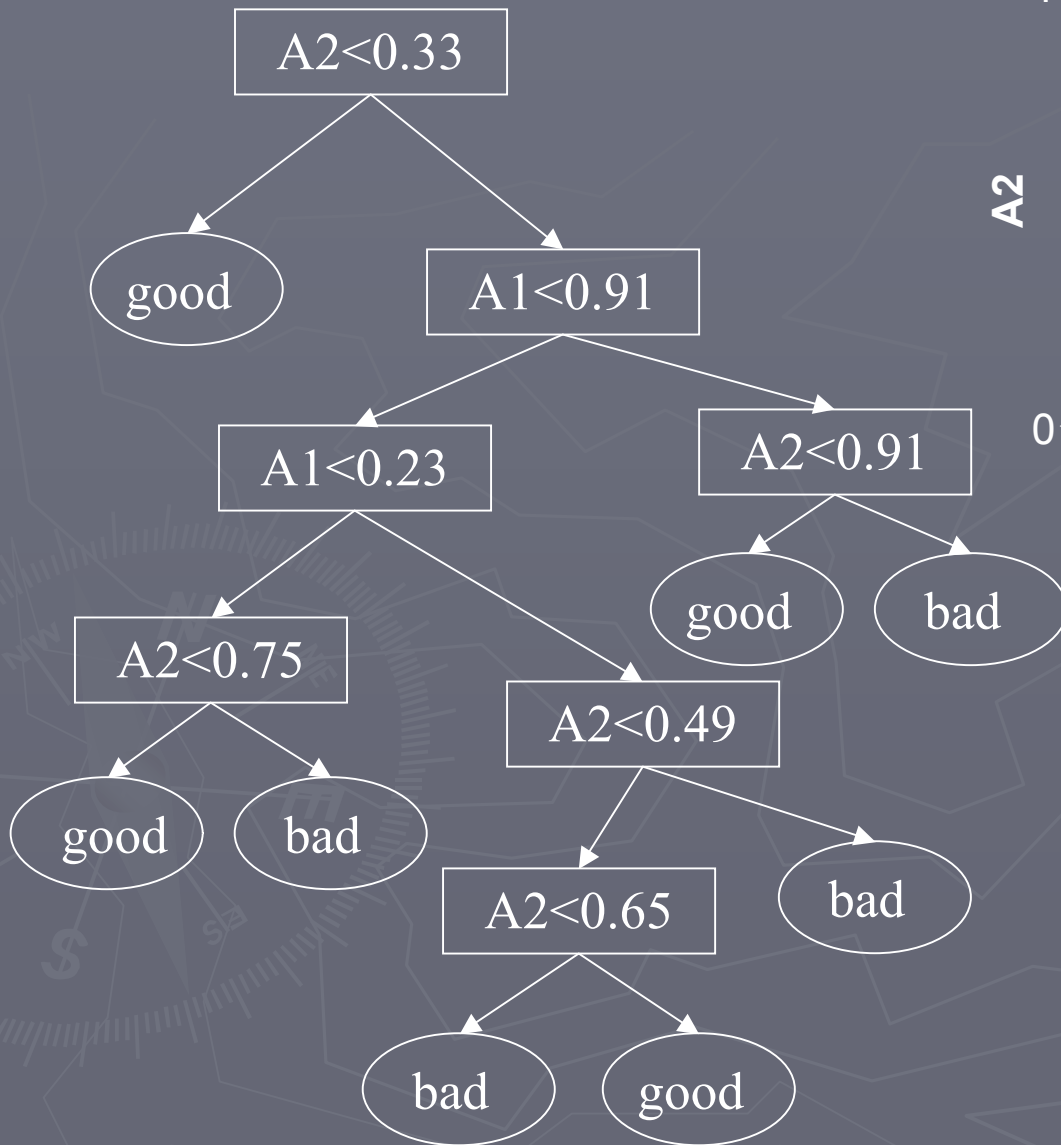
$A2 < 0,33$



# Tree growing

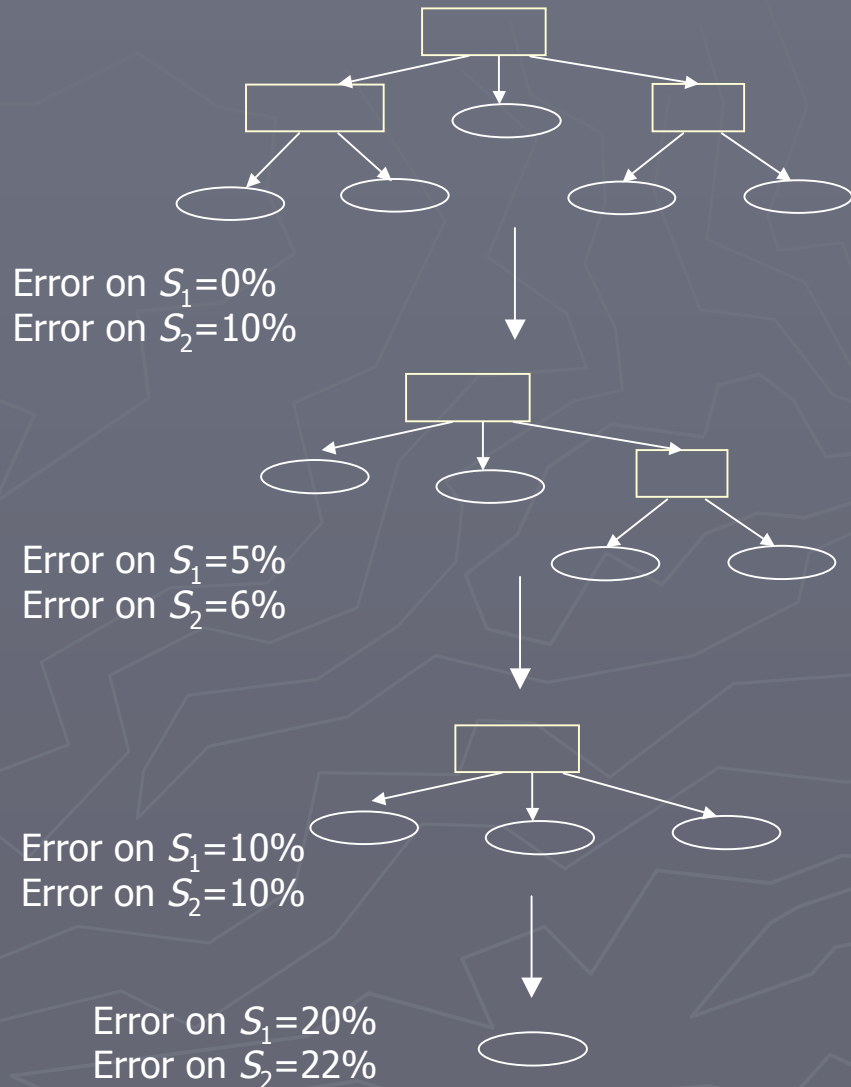


# Tree growing

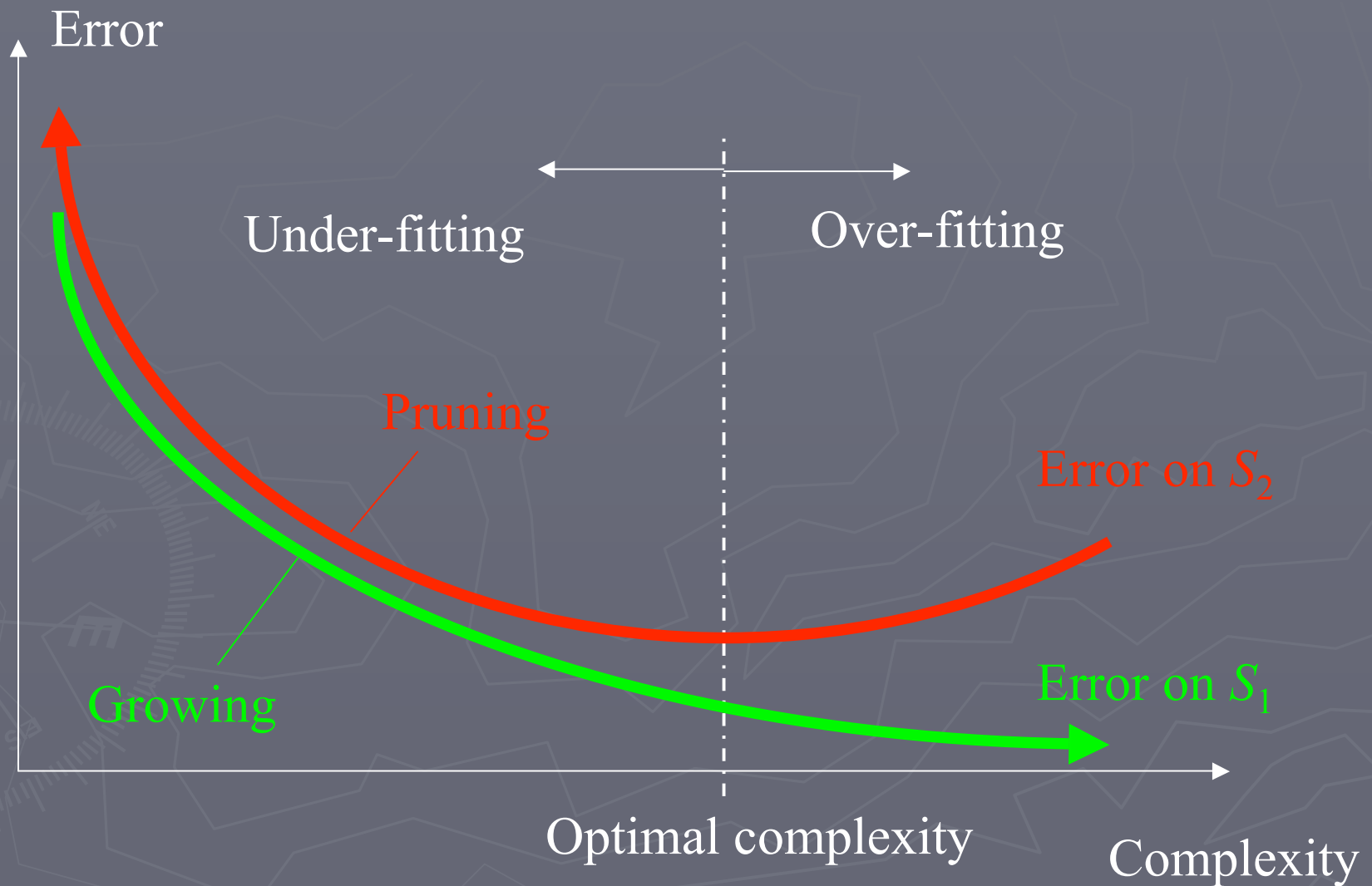


# Tree pruning

- ▶ Too avoid overfitting
- ▶ Grow a too large tree and then, reduce it by removing irrelevant parts
- ▶ In practice, pruning is very fast



# Tree pruning



# Interpretability and attribute selection

## ► Interpretability

- Intrinsically, a decision tree is highly interpretable
- A tree may be converted into a set of “if...then” rules.

## ► Attribute selection

- If some attributes are not useful for classification, they will not be selected in the (pruned) tree
- Of practical importance, if measuring the value of an attribute is costly (e.g. medical diagnosis)
- Decision trees are often used as a pre-processing for other learning algorithms that suffer more when there are irrelevant variables

# Demo

▶ Database of Titanic passengers:

- <http://www.cs.toronto.edu/~delve/data/titanic/desc.html>

▶ Java applet

- <http://www.montefiore.ulg.ac.be/~geurts/dtapplet/dtexplication.html>

# In practice

- ▶ Very fast and scalable method (able to handle a very large number of input variables and learning sample cases)
- ▶ Provide directly interpretable models and give an idea of the relevance of attributes.
- ▶ Often less accurate than other algorithms like artificial neural networks